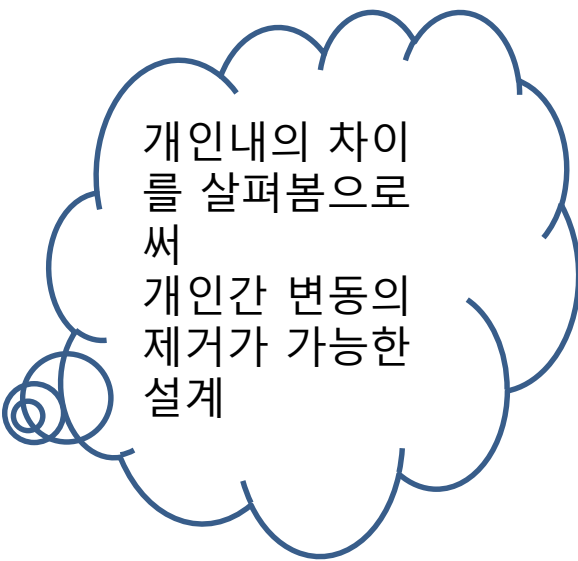


집단 비교 방법의 예제

Example : Dietary intake

- Mean daily dietary intake(kJ) over 10 pre-menstrual and 10 post-menstrual days (Manocha et al., 1986)

subject	pre_m	post_m	diff
1	5260	3910	1350
2	5470	4220	1250
3	5640	3885	1755
4	6180	5160	1020
5	6390	5645	745
6	6515	4680	1835
7	6805	5265	1540
8	7515	5975	1540
9	7515	6790	725
10	8230	6900	1330
11	8770	7335	1435
mean	6753.64	5433.18	1320.45
s.d.	1142.12	1216.83	366.75

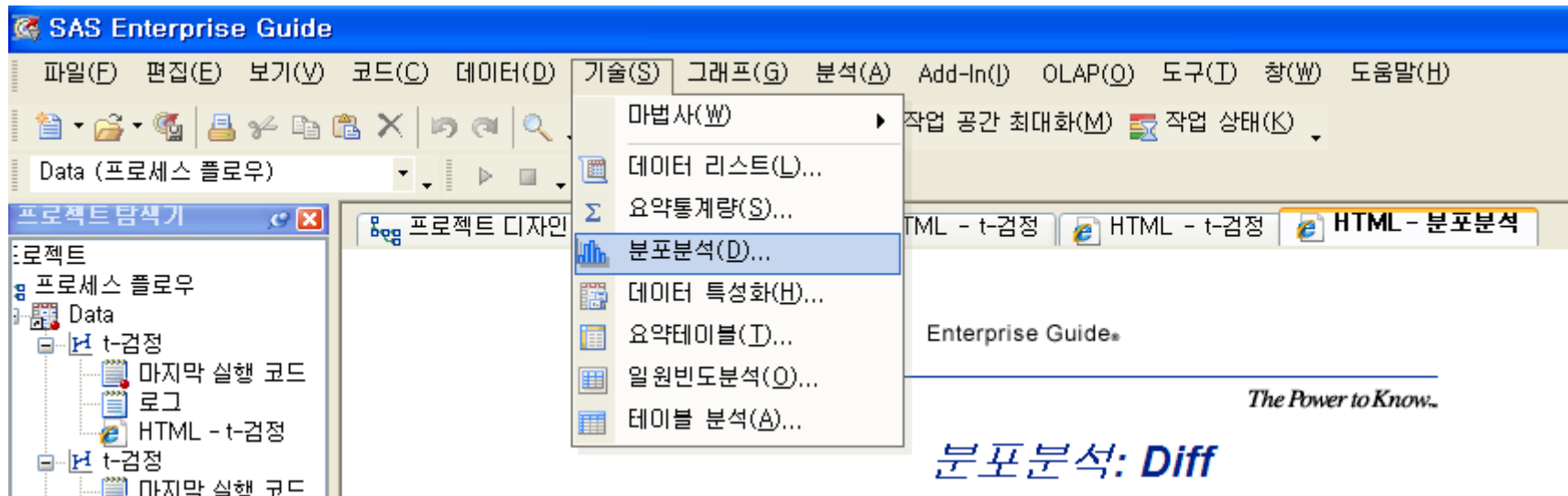


개인내의 차이를 살펴봄으로써
개인간 변동의 제거가 가능한
설계

SAS : Normaility

- Test of Normality

- Plot : Normal Probability Plot(NPP), Quantile-Quantile plot(Q-Q plot)
- Test : Shapiro-Wilk's test, Kolmogorov-Smirnov's Test / Chi-square test



DATA에 대한 분포분석

작업 역할

분포

요약

정규

로그 정규

지수

와이블

베타

감마

커널

도표

모양

인셋

테이블

제목

작업 역할

할당할 변수(A):

이름

subject

Pre_M

Post_M

Diff

작업 역할(E):

분석변수

Diff

그룹 분석변수

빈도변수 (제한: 1개)

상대 가중값 변수 (제한: 1개)

분류변수 (제한: 2개)

DATA에 대한 분포분석

작업 역할

분포

요약

정규

로그 정규

지수

와이블

베타

감마

커널

도표

모양

인셋

테이블

제목

테이블

☒ 기본 신뢰구간
 ☒ 기본 속도
 ☒ 위치모수에 대한 검정
 ☐ 극 관측치
 ☐ 극 값
 ☐ 빈도
 ☐ 최빈값
 ☐ 적률
 ☐ 분위수
 ☐ 척도모수의 로버스트 속도
 ☒ 정규성 검정
 ☐ 절사 평균
 ☐ 원저화 평균

– 정규성 검정과 정규분포에 대한 적합성 검정

정규성 검정				
검정	통계량		p-값	
Shapiro-Wilk	W	0.937371	Pr < W	0.4901
Kolmogorov-Smirnov	D	0.151102	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.0433	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.296907	Pr > A-Sq	>0.2500

정규 분포에 대한 적합도 검정				
검정	통계량		p-값	
Kolmogorov-Smirnov	D	0.15110175	Pr > D	>0.150
Cramer-von Mises	W-Sq	0.04329964	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.29690723	Pr > A-Sq	>0.250

SAS : Paired T-test

SAS Enterprise Guide

파일(F) 편집(E) 보기(V) 코드(C) 데이터(D) 기술(S) 그래프(G) 분석(A) Add-In(I) OLAP(O) 도구(T) 창(W) 도움말(H)

프로젝트 디자인(B) 작업 공간 최대화(M) 작업 상태(K)

Data (프로세스 플로우)

프로젝트 탐색기

- 프로젝트
 - 프로세스 플로우
 - Data

	subject	Pre_M	Post_M
1	1	5260	3910
2	2	5470	4220
3	3	5640	3885
4	4	6180	5160
5	5	6390	5645
6			
7			
8			
9			
10			
11			

기술(S) 그래프(G) 분석(A) Add-In(I) OLAP(O) 도구(T) 창(W) 도움말(H)

프로젝트 디자인(B) Data (읽기 전용)

분석(A)

- t-검정(T)...
- 일원분산분석(O)...
- 비모수적 일원분산분석(N)...
- 선형모형(L)...
- 혼합모형(M)...

회귀(R) 다변량(M) 생존분석(S) 공정능력분석(B) 관리도(O) 파레토 그래프(P)... 시계열(T) 모형 스코어링(L)...

The Power to Know.

- 모수적 방법 : Paired t-test

Assumption : Normal Distribution

- 1) direct
- 2) using difference

					Statistics					
Difference	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err	Minimu m	Maximu m
Post_M - Pre_M	11	-1567	-1320	-1074	256.25	366.75	643.61	110.58	-1835	-725

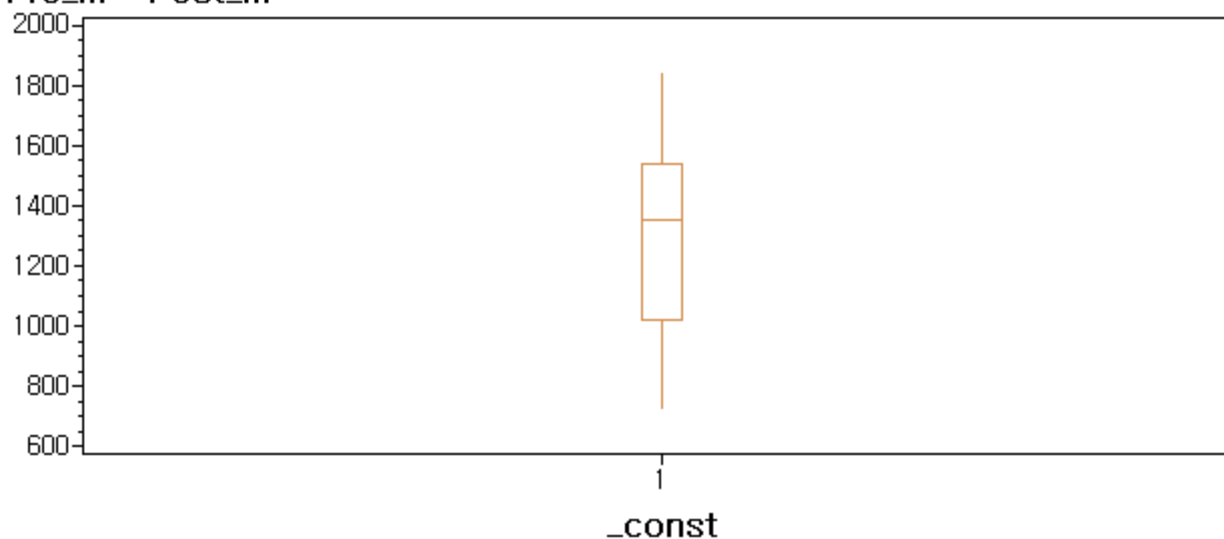
T-Tests			
Difference	DF	t Value	Pr > t
Post_M - Pre_M	10	-11.94	<.0001

$$t = \frac{\bar{d} - H_0}{se(\bar{d})} = \frac{\bar{d} - H_0}{s.d.(d) / \sqrt{n}} = \frac{-1320.5 - 0}{366.7 / \sqrt{11}} = \frac{-1320.5}{110.6} = -11.94$$

Statistics										
Variable	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err	Minimum	Maximum
Diff	11	1074.1	1320.5	1566.8	256.25	366.75	643.61	110.58	725	1835

T-Tests			
Variable	DF	t Value	Pr > t
Diff	10	11.94	<.0001

Pre_M - Post_M



Example : Nonparametric

- 비모수적 방법 : 대응자료분석 및 one sample sign test
 - Assumption : distribution-free
 - 1) One sample sign test
 - 2) Wilcoxon matched pairs signed rank sum test
- **One sample sign test**
 - 귀무가설 : no differences 혹은 부위(+/-)가 동일
[참고 : 11명 자료 중 5.5명이 증가하고 감소하면 차이가 없다는 의미로 검정]
 - 검정통계량
$$\frac{|11 - 5.5| - 0.5}{\sqrt{11 \times 0.5 \times (1 - 0.5)}} = 3.02 \sim \text{app. Normal}$$
 - p-값 = 0.003 (기각)

- **Wilcoxon matched pairs signed rank sum test**

- 기본 개념 : 증가군의 순위와 감소군의 순위를 비교하여 검정하는 방법
- 과정
 - 1) 차이 계산 : 차이가 '0'인 자료 제외
 - 2) 절대값 차이의 순위값 부여 : 차이가 동일한 경우 평균순위부여
 - 3) 순위값에 차이의 부호 부여
- 검정통계량 (동점자료가 없는 경우)

$$T_1 = \frac{\sum R_i}{\sqrt{\sum R_i^2}}$$

$$T_1 = \frac{\sum R_i}{\sqrt{\frac{n(n+1)(2n+1)}{6}}}$$

(+ 를 이용하는 경우) $T^+ = \sum_1^n \psi(D_i) \cdot R_i^+$ where $\psi(D_i) = 1, D_i > 0$ and $0, D_i \leq 0$

- 기각기준

- 일반형, 동점이 없거나 적은 경우 : 정규분포로 가정하여 검정
- T^+ 를 이용하는 경우 : 별도의 분위표를 이용

Example : Hodgkins disease

- Numbers of T4 and T8 cells/mm³ in blood samples in remission from Hodgkins disease (Shapiro et al., 1986)

id	Gr	T4	T8	T4-T8	Abs(T4-T8)	Rank
1	Hodgkins	396	836	-440	440	13
2	Hodgkins	568	978	-410	410	12
3	Hodgkins	1212	1678	-466	466	14
4	Hodgkins	171	212	-41	41	4
5	Hodgkins	554	670	-116	116	7
6	Hodgkins	1104	1335	-231	231	10
7	Hodgkins	257	272	-15	15	2
8	Hodgkins	435	446	-11	11	1
9	Hodgkins	295	262	33	33	3
10	Hodgkins	397	340	57	57	6
11	Hodgkins	288	236	52	52	5
12	Hodgkins	1004	786	218	218	9
13	Hodgkins	431	311	120	120	8
14	Hodgkins	795	449	346	346	11
15	Hodgkins	1621	811	810	810	18
16	Hodgkins	1378	686	692	692	17
17	Hodgkins	902	412	490	490	15
18	Hodgkins	958	286	672	672	16
19	Hodgkins	1283	336	947	947	19
20	Hodgkins	2415	936	1479	1479	20

순위합
=63

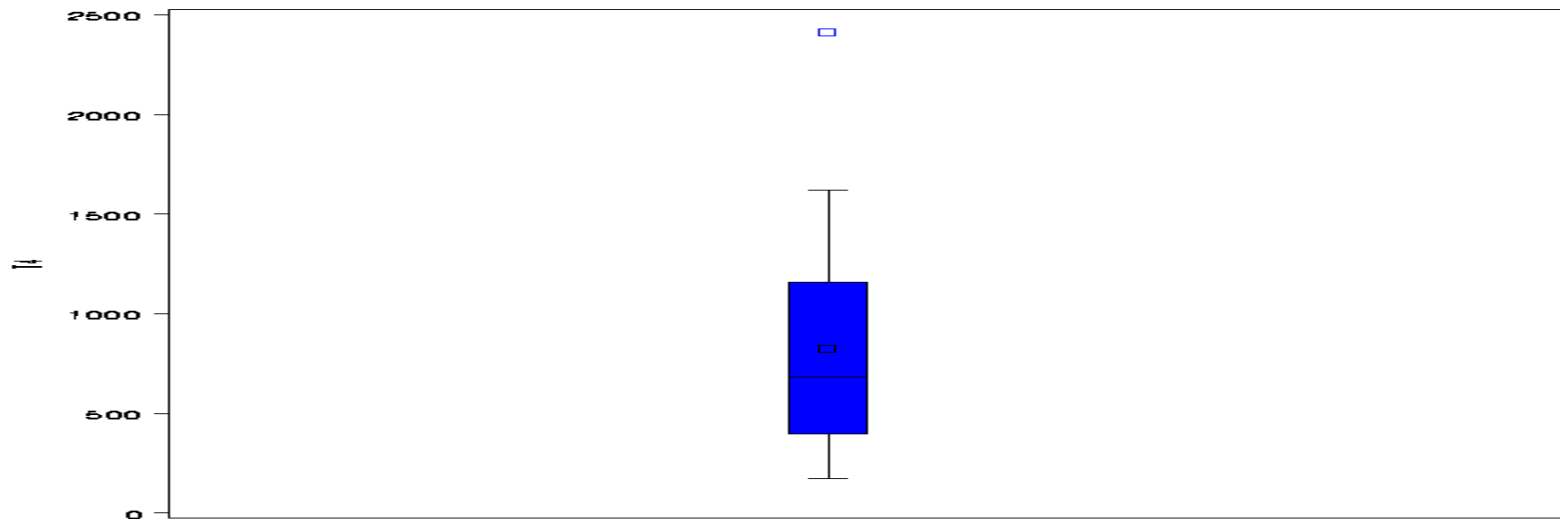
순위합
=147

SAS : Wilcoxon signed rank test

- 정규성 검정 : T4 자료에 대한 검토

정규성 검정

검정	통계량	p-값
Shapiro-Wilk	W 0.891445	Pr < W 0.0286
Kolmogorov-Smirnov	D 0.173853	Pr > D 0.1116
Cramer-von Mises	W-Sq 0.093205	Pr > W-Sq 0.1326
Anderson-Darling	A-Sq 0.650398	Pr > A-Sq 0.0804

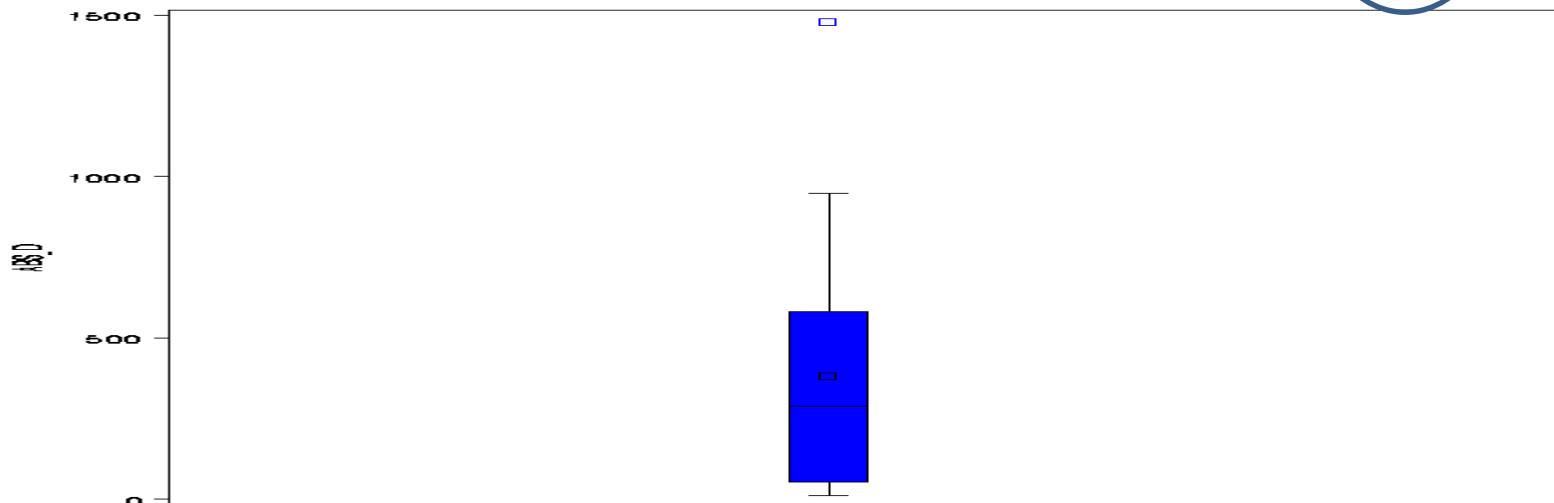


- 정규성 검정 : 절대값 차이 자료에 대한 검토

정규성 검정

검정	통계량	p-값
Shapiro-Wilk	W 0.861128	Pr < W 0.0082
Kolmogorov-Smirnov	D 0.167759	Pr > D 0.1438
Cramer-von Mises	W-Sq 0.123986	Pr > W-Sq 0.0488
Anderson-Darling	A-Sq 0.833335	Pr > A-Sq 0.0258

정규성
의심



The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable abs_D
Classified by Variable gr

gr	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1	8	63.0	84.0	12.961481	7.8750
2	12	147.0	126.0	12.961481	12.2500

Wilcoxon Two-Sample Test

Statistic 63.0000

Normal Approximation

Z -1.5816
One-Sided Pr < Z 0.0569
Two-Sided Pr > |Z| 0.1137

t Approximation

One-Sided Pr < Z 0.0651
Two-Sided Pr > |Z| 0.1302

Z includes a continuity correction of 0.5.

Kruskal-Wallis Test

Chi-Square 2.6250
DF 1
Pr > Chi-Square 0.1052

양측검정으로 판단
하면 귀무가설 채택
하므로 차이가 없음

[참고 1] outlier (skewed data)의 해결 : 정규분포를 이용하는 경우

1) 제거 2) 자료변환

[참고 2] 모수방법 대신 비모수방법 이용 검토

Example : Energy expenditure

- 24hour total energy expenditure (MJ/day) in groups of lean and obese women (Prentice et al., 1986)

Lean	Obese
6.13	8.79
7.05	9.19
7.48	9.21
7.48	9.68
7.53	9.69
7.58	9.97
7.90	11.51
8.08	11.85
8.09	12.79
8.11	
8.40	
10.15	
10.88	

SAS : Two Independent Groups

SAS Enterprise Guide

파일(F) 편집(E) 보기(V) 코드(C) 데이터(D) 기술(S) 그래프(G) 분석(A) Add-In(I) OLAP(O) 도구(T) 창(W) 도움말(H)

프로젝트 탐색기

프로젝트
프로세스 플로우
Data

Data (프로세스 플로우)

프로젝트 디자인

	Gr	Energy
1	Lean	6.13
2	Lean	7.05
3	Lean	7.48
4	Lean	7.48
5	Lean	7.53

분석(A)

- 분산분석(A)
 - t-검정(T)...
 - 일원분산분석(O)...
 - 비모수적 일원분산분석(N)...
 - 선형모형(L)...
 - 혼합모형(M)...
- 회귀(B)
- 다변량(M)
- 생존분석(S)
- 공정능력분석(B)
- 관리도(O)
- 파레토 그래프(P)...
- 시계열(T)

DATA에 대한 t-검정

t-검정 유형

작업 역할
분석
도표
제목

t-검정 유형 선택:

- ☒ 미표본(I)
- ☐ 쌍체(P)
- ☐ 일표본(O)

DATA에 대한 t-검정

t 검정 유형

작업 역할

분석

도표

제목

작업 역할

할당할 변수(A):

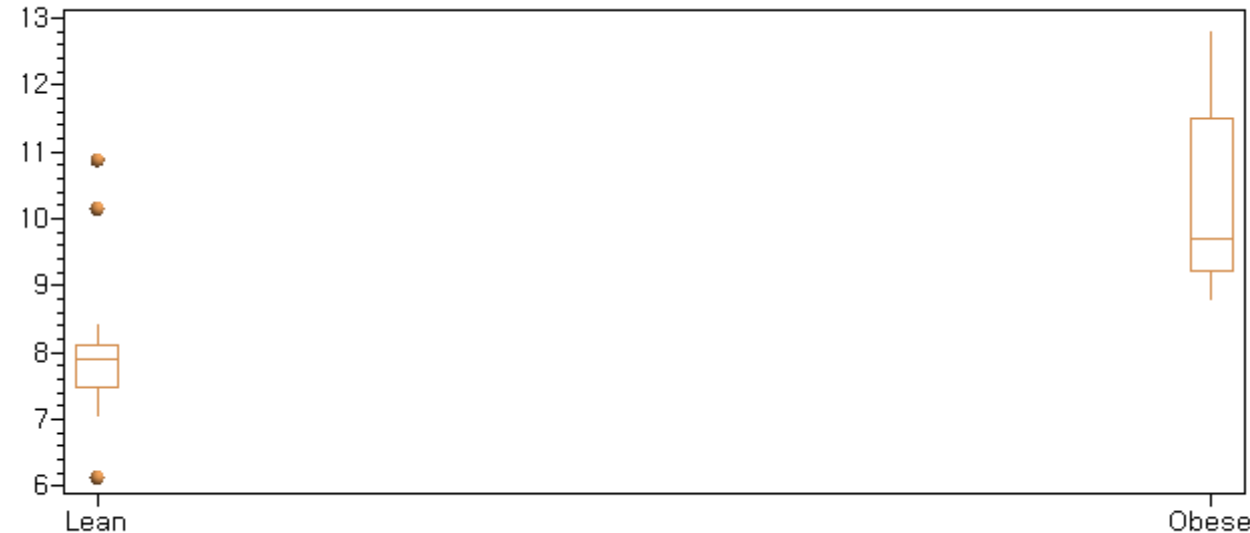
이름

- △ Gr
- ⑫ Energy
- △ C
- △ D
- △ E
- △ F

작업 역할(E):

- 그룹화 (제한: 1개)
 - △ Gr
- 분석변수
 - ⑫ Energy
- 그룹 분석변수
- 빈도변수 (제한: 1개)
- 상대 가중값 변수 (제한: 1개)

Energy



Gr

Equality of Variances					
Variable	Method	Num DF	Den DF	F Value	Pr > F
Energy	Folded F	8	12	1.27	0.6797

등분산
가정 만족

Statistics										
Variable	Gr	N	Lower CL Mean	Mean	Upper CL Mean	Lower CL Std Dev	Std Dev	Upper CL Std Dev	Std Err	Min Max
Energy	Lean	13	7.318	8.0662	8.8143	0.8878	1.2381	2.0437	0.3434	6.13 10.88
Energy	Obese	9	9.2233	10.298	11.372	0.9442	1.3979	2.678	0.466	8.79 12.79
Energy	Diff (1-2)		-3.411	-2.232	-1.052	0.9979	1.3043	1.8836	0.5656	

T-Tests						
Variable	Method	Variances	DF	t Value	Pr > t	
Energy	Pooled	Equal	20	-3.95	0.0008	
Energy	Satterthwaite	Unequal	15.9	-3.86	0.0014	

유의한
차이 존재

살찐 여성과 날씬한 여성의 에너지 소비량에 대한 차이를 유의수준 5%수준에서 살펴보면, 통계적으로 유의한 차이를 보여주고 있으며, 차이는 2MJ정도임.

Example : Nonparametric

● 비모수적 방법 : 독립 표본(집단)에 대한 검정

– Assumption : distribution-free, independent groups

1) Mann-Whitney test : U [참고 : $U = n_1 n_2 - \frac{1}{2} n_1 (n_1 + 1) - T$]

2) Whitney-Wilcoxon test : T (표본수가 작은 그룹의 순위합)

– 특징 : 왜도가 심하거나 특이값이 존재하는 경우에 유용

– 분석과정

1) 집단 구분없이 자료를 크기순으로 배열하여 순위 부여

2) 집단별 순위합 계산

– 검정통계량 : $T = \sum R(X_i)$ (동점이 없는 경우)

(동점이 많은 경우)

$$T_1 = \frac{T - E(T)}{\sqrt{\text{Var}(T)}} = \frac{T - n \left(\frac{N+1}{2} \right)}{\sqrt{\frac{n \cdot m}{N(N-1)} \sum_1^N R_i^2 - \frac{n \cdot m (N+1)^2}{4(N-1)}}$$

– 기각기준 : 분위표/정규분포표 이용

SAS : Mann-Whitney

The screenshot displays the SAS software interface. The main menu bar includes '기술(S)', '그래프(G)', '분석(A)', 'Add-In(I)', 'OLAP(O)', '도구(T)', '창(W)', and '도움말(H)'. The '분석(A)' menu is open, showing a list of statistical tests. The '비모수적 일원분산분석(N)...' option is highlighted. Below this, a sub-menu is visible with the text 'used for ties.' and 'Sample Test'. The 'DATA에 대한 비모수적 일원분산분석' dialog box is open, showing the '작업 역할' (Job Role) section. The '작업 역할' section is divided into '작업 역할' (Job Role) and '작업 역할(E):' (Job Role (E)). The '작업 역할' section lists the following variables: Gr, Energy, C, D, and E. The '작업 역할(E):' section lists the following variables: 종속변수 (Dependent Variable), 독립변수 (제한: 1개) (Independent Variable (Limit: 1)), Gr, 그룹 분석변수 (Group Analysis Variable), and 빈도변수 (제한: 1개) (Frequency Variable (Limit: 1)).

기술(S) 그래프(G) 분석(A) Add-In(I) OLAP(O) 도구(T) 창(W) 도움말(H)

분석(A) 메뉴:

- 분산분석(A)
 - t-검정(T)...
 - 일원분산분석(O)...
 - 비모수적 일원분산분석(N)...**
 - 선형모형(L)...
 - 혼합모형(M)...
- 회귀(R)
- 다변량(M)
- 생존분석(S)
- 공정능력분석(B)
- 관리도(Q)
- 파레토 그래프(P)...
- 시계열(T)
- 모형 스코어링(L)...

비모수적 일원분산분석(N)...

used for ties.

Sample Test

150.0000

DATA에 대한 비모수적 일원분산분석

작업 역할

분석
정확 p-값
결과
제목

작업 역할

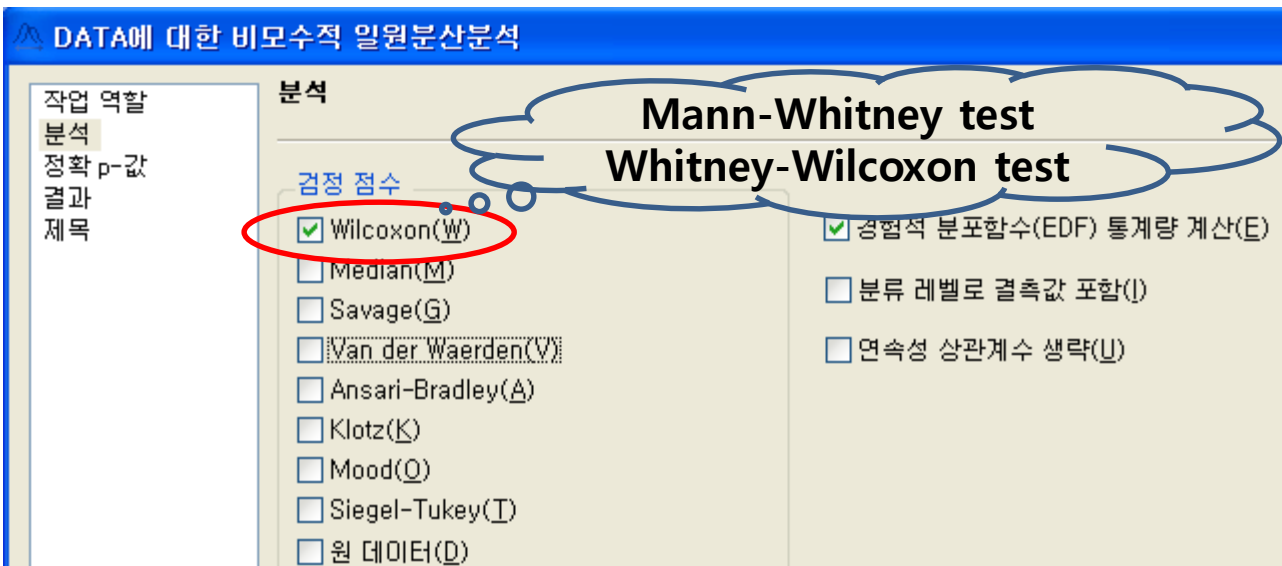
할당할 변수(A):

이름

- Gr
- Energy
- C
- D
- E

작업 역할(E):

- 종속변수
- Energy
- 독립변수 (제한: 1개)
- Gr
- 그룹 분석변수
- 빈도변수 (제한: 1개)



Wilcoxon Scores (Rank Sums) for Variable Energy Classified by Variable Gr					
Gr	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Lean	13	103.0	149.50	14.970751	7.923077
Obese	9	150.0	103.50	14.970751	16.666667
Average scores were used for ties.					

Wilcoxon Two-Sample Test

Statistic (S) 150.0000

Normal Approximation

Z 3.0727

One-Sided Pr > Z 0.0011

Two-Sided Pr > |Z| 0.0021

t Approximation

One-Sided Pr > Z 0.0029

Two-Sided Pr > |Z| 0.0058

Exact Test

One-Sided Pr >= S 5.287E-04

Two-Sided Pr >= |S - Mean| 0.0010

Z includes a continuity correction of 0.5.

유의한
차이 존재

Kruskal-Wallis Test

Chi-Square 9.6476

DF 1

Pr > Chi-Square 0.0019

Example : Cardiac Bypass

- Red Cell folate levels(ug/l) in three groups of cardiac bypass patients after 24h ventilation(Amess et al., 1978)
 - Gr I : 50% nitrous oxide and 50% oxygen mixture(24h)
 - Gr II : 50% nitrous oxide and 50% oxygen mixture(operation)
 - Gr III : no nitrous oxide and 30-50% oxygen (24h)

Gr1	Gr2	Gr3
243	206	241
251	210	258
275	226	270
291	249	293
347	255	328
354	273	
380	285	
392	295	
	309	

SAS : ANOVA

코드(C) 데이터(D) 기술(S) 그래프(G) 분석(A) Add-In(I) OLAP(O) 도구(T) 창(W) 도움말(H)

프로젝트 디자인

	Gr	Folate
1	Gr1	243
2	Gr1	251
3	Gr1	275
4	Gr1	291
5	Gr1	347
6	Gr1	354
7	Gr1	380

분석(A) 메뉴:

- 분산분석(A)
 - t-검정(T)...
 - 일원분산분석(O)...
 - 비모수적 일원분산분석(N)...
 - 선형모형(L)...
 - 혼합모형(M)...
- 회귀(B)
- 다변량(M)
- 생존분석(S)
- 공정능력분석(B)
- 관리도(O)
- 파레토 그래프(P)...
- 시계열(T)
- 모형 스코어링(L)...

DATA에 대한 일원분산분석

작업 역할

검정
평균
평균 비교
통계량 계산
도표
결과
제목

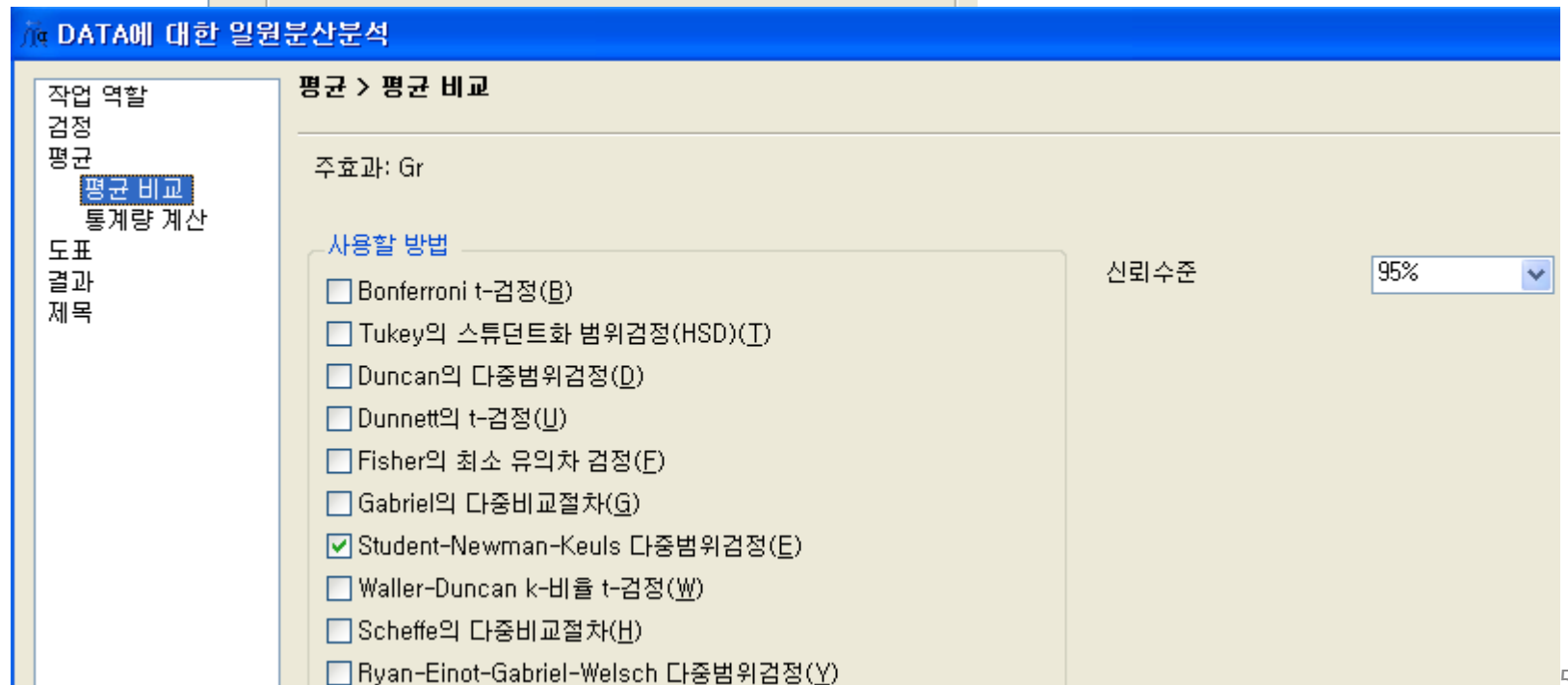
작업 역할

할당할 변수(A):

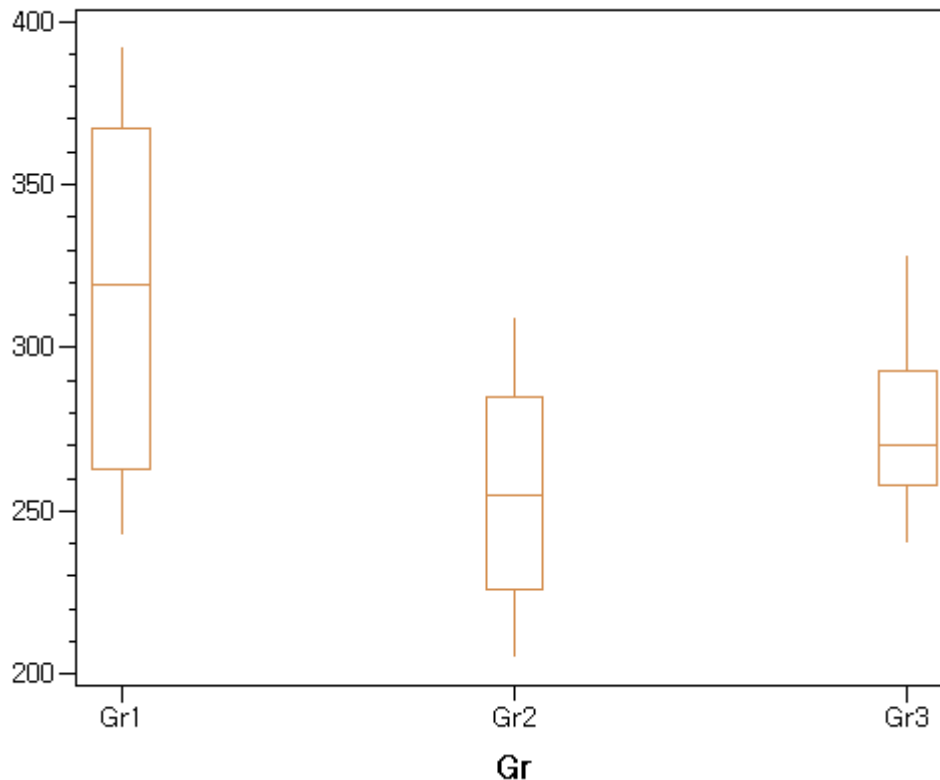
- 이름
- Gr
- Folate
- C
- D

작업 역할(E):

- 종속변수
- Folate
- 독립변수 (제한: 1개)
- Gr
- 그룹 분석변수



Folate



Level of Gr	N	Folate	
		Mean	Std Dev
Gr1	8	316.625000	58.7170880
Gr2	9	256.444444	37.1217965
Gr3	5	278.000000	33.7564809

분산 동질성 검정

1) Levene : 표준검정

2) Bartlett : 우도비검정
수정, 정규분포가정

3) Welch : 등분산 기각
시 사용

Levene's Test for Homogeneity of Folate Variance ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Gr	2	18765720	9382860	4.14	0.0321
Error	19	43019786	2264199		

Bartlett's Test for Homogeneity of Folate Variance

Source	DF	Chi-Square	Pr > ChiSq
Gr	2	2.0951	0.3508

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	15515.76641	7757.88321	3.71	0.0436
Error	19	39716.09722	2090.32091		
Corrected Total	21	55231.86364			

R-Square	Coeff Var	Root MSE	Folate Mean
0.280921	16.14252	45.72003	283.2273

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Gr	2	15515.76641	7757.88321	3.71	0.0436

심장 보조관 수술환자에 대한 질소와 산소 공급 수준에 따른 폴산염 수준의 차이가 존재하는 지 분산분석으로 살펴본 결과 유의수준 5%수준에서 유의미한 차이가 존재하는 것으로 나타남.

- Multiple comparisons : S-N-K

- This test controls the Type I experimentwise error rate under the complete null hypothesis but not under partial null hypotheses.

Alpha	0.05	
Error Degrees of Freedom	19	
Error Mean Square	2090.321	
Harmonic Mean of Cell Sizes	6.878981	
Number of Means	2	3
Critical Range	51.598112	62.62823

평균 순위차이
= $j-i+1$

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	Gr
A	316.63	8	Gr1
A	278.00	5	Gr3
A	256.44	9	Gr2

그룹간 차이는
존재하지 않음

- Multiple comparisons : Duncan

- This test controls the Type I comparisonwise error rate, not the experimentwise error rate.

Alpha	0.05	
Error Degrees of Freedom	19	
Error Mean Square	2090.321	
Harmonic Mean of Cell Sizes	6.878981	
Number of Means	2	3
Critical Range	51.60	54.15

평균 순위차이
= $j-i+1$

Means with the same letter are not significantly different.

SNK Grouping	Mean	N	Gr
A	316.63	8	Gr1
B A	278.00	5	Gr3
B	256.44	9	Gr2

그룹간 차이는
존재

Example : Nonparametric

● 비모수적 방법 : Kruskal-Wallis 검정

- Assumption : distribution-free, independent groups
- 특징 : Mann-Whitney 검정의 확장
- 분석과정

1) 집단 구분없이 자료를 크기순으로 배열하여 순위 부여

2) 집단별 평균순위 계산

- 검정통계량 :

$$T_1 = \left[\sum_1^k \frac{R_i^2}{n_i} - \frac{N(N+1)^2}{4} \right] / \left[\frac{1}{N-1} \left\{ \sum_1^k \sum_1^{n_i} R_{ij}^2 - \frac{N(N+1)^2}{4} \right\} \right]$$

(동점이 없는 경우)

$$T = \frac{12 \sum_1^k n_i (\bar{R}_i - \bar{R})^2}{N(N+1)} = \frac{12}{N(N+1)} \sum_1^k \frac{R_i^2}{n_i} - 3(N+1)$$

- 기각기준 : 카이제곱분포표 이용

SAS : Kruskal-Wallis

Wilcoxon Scores (Rank Sums) for Variable Folate Classified by Variable Gr

Gr	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Gr1	8	120.0	92.00	14.651507	15.000000
Gr2	9	77.0	103.50	14.974979	8.555556
Gr3	5	56.0	57.50	12.763881	11.200000

Kruskal-Wallis Test

Chi-Square 4.1852

DF 2

Pr > Chi-Square 0.1234

집단간 차이가 존재하는지에 대해 비모수방법으로 살펴본 결과, 5%의 유의수준에서 유의한 차이가 없는 것으로 나타남.

ANOVA : Two-way

- 두 요인 A(i), B(j)의 영향에 대한 연구 : 반복 실험을 가정

	수준B1	수준B2	...	수준Bj
수준A1	y_{111} ...	y_{121}	y_{1J1} ...
	y_{11n}	y_{12n}		y_{1Jn}
	$\bar{y}_{11.}$	$\bar{y}_{12.}$...	$\bar{y}_{1J.}$
...
수준Ai	y_{I11} ...	y_{I21}	y_{IJ1} ...
	y_{I1n}	y_{I2n}		y_{IJn}
	$\bar{y}_{I1.}$	$\bar{y}_{I2.}$...	$\bar{y}_{IJ.}$

$$\bar{y}_{ij.} = \sum_k y_{ijk} / n$$

$$\bar{y}_{i..} = \sum_j \sum_k y_{ijk} / (Jn)$$

$$\bar{y}_{.j.} = \sum_i \sum_k y_{ijk} / (In)$$

$$\bar{y}_{...} = \sum_i \sum_j \sum_k y_{ijk} / (IJn)$$

- 분산분석모형 : α, β 는 주효과, $(\alpha\beta)$ 는 교호작용효과

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad \text{where } \varepsilon_{ijk} \sim N(0, \sigma^2)$$

- 가정 : 오차항~독립, 정규분포, 등분산 가정
- 효과에 대한 조건 : (모수인자 ; fixed factor)

$$\sum_i \alpha_i = 0, \sum_j \beta_j = 0, \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$$

- 분석 관점

- 주효과(main effects) 에 대한 분석
- 교호작용효과(interaction effects)에 대한 분석

- 분산분석표

요인	제곱합(SS)	자유도(df)	평균제곱(MS)	F-ratio
A	SSA	dfA=I-1	MSA	MSA/MSE
B	SSB	dfB=J-1	MSB	MSB/MSE
AXB	SSAXB	dfAxB=(I-1)(J-1)	MSAxB	MSAxB/MSE
Error	SSE	dfE=IxJx(n-1)	MSE	
Total	SST	(IxJxn)-1		

– 여기서

$$SST = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2 \quad SSA = (Jn) \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2$$

$$SSB = (In) \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 \quad SSAxB = n \sum_i \sum_j (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$SSE = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2$$

$$MSk = \frac{SSk}{dfk}, \quad k = A, B, AxB$$

• 가설의 기각 판단

- 일반적인 가설 검정의 원리와 동일 : F-분포 이용
- 검정 과정
 - ① 교호작용효과에 대한 검정 : 유의하지 않으면 주효과 검정
 - ② 주효과에 대한 검정
 - ③ 다중비교 및 사후분석으로 동일집단 분류

[참고]

1. 분산분석 모형에서 교호작용효과가 유의하면 주효과분석은 의미가 없으므로 교호작용효과가 유의하지 않은 경우에 주효과에 대한 검정을 수행함.
2. 분산분석모형을 표현할 때 교호작용이 유의하면 주효과는 모형에 반드시 포함시켜 표현해야 함.
3. 반복이 없는 경우는 교호작용효과가 존재하지 않음.
4. 사후분석 및 다중비교는 귀무가설이 기각된 경우에 수행하며, 다중비교방법에 따라 집단간 차이유무의 결과가 다르게 나타남.

Example : 혈액 칼슘 자료

- 호르몬 처리(3종)와 성별(2종)에 따른 혈액 칼슘값의 차이가 존재하는 지에 대한 자료(교재 p.30)
 - 남녀별 15명을 랜덤하게 배정하여 실험

	Tr1	Tr2	Tr3
Male	16.87	19.07	32.05
	16.18	18.77	28.71
	17.12	17.63	34.65
	16.83	16.99	28.79
	17.19	18.04	24.46
Female	15.86	17.20	30.54
	14.92	17.64	32.41
	15.63	17.89	28.97
	15.24	16.78	28.46
	14.80	16.72	29.65

SAS : Two-way ANOVA

SAS Enterprise Guide

파일(F) 편집(E) 보기(V) 코드(C) 데이터(D) 기술(S) 그래프(G) 분석(A) Add-In(I) OLAP(O) 도구(T) 창(W) 도움말(H)

프로젝트 탐색기

프로젝트

프로세스 플로우

Data

프로젝트 디자인

	Gender	Trt
1	Male	Tr1
2	Male	Tr1
3	Male	Tr1

분석(A)

- 분산분석(A)
 - t-검정(T)...
 - 일원분산분석(O)...
 - 비모수적 일원분산분석(N)...
 - 선형모형(L)...
 - 혼합모형(M)...
- 회귀(B)
- 다변량(M)
- 생존분석(S)
- 공정능력분석(B)
- 관리도(Q)

DATA에 대한 선형모형

작업 역할

모형

모형 옵션

고급 옵션

Post Hoc 검정

최소제곱

산술

도표

평균

예측값

잔차

영향력

예측값

제목

작업 역할

할당할 변수(A):

이름

Gender

Trt

Ca

작업 역할(E):

- 종속변수 (제한: 1개)
 - Ca
- 양적 변수
- 분류변수
 - Trt
 - Gender
- 그룹 분석변수
- 빈도변수 (제한: 1개)
- 상대 가중값 변수 (제한: 1개)

DATA에 대한 선형모형

작업 역할

모형

모형 옵션

고급 옵션

Post Hoc 검정

최소제곱

산술

도표

평균

예측값

모형

분류변수 및 양적변수(V):

Gender

Trt

주효과(M)

교차(Q)

지분효과(N)

효과(E):

Gender

Trt

Gender*Trt

DATA에 대한 선형모형

작업 역할

모형

모형 옵션

고급 옵션

Post Hoc 검정

최소제곱

산술

도표

평균

예측값

잔차

영향력

예측값

제목

모형 옵션

가설 검정

☐ 절편에 대한 검정 표시(I)

표시할 제공합

☒ Type I(I)

☐ Type II(Y)

☒ Type III(P)

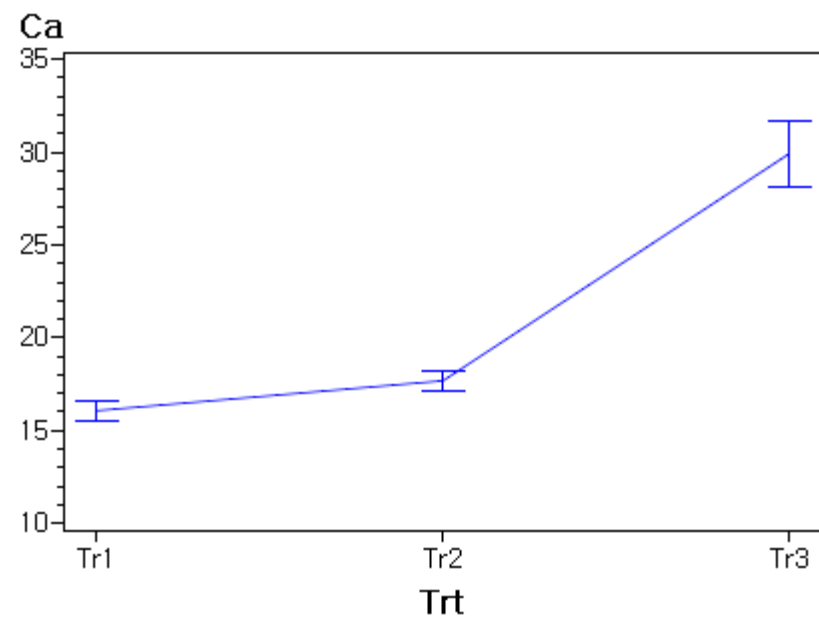
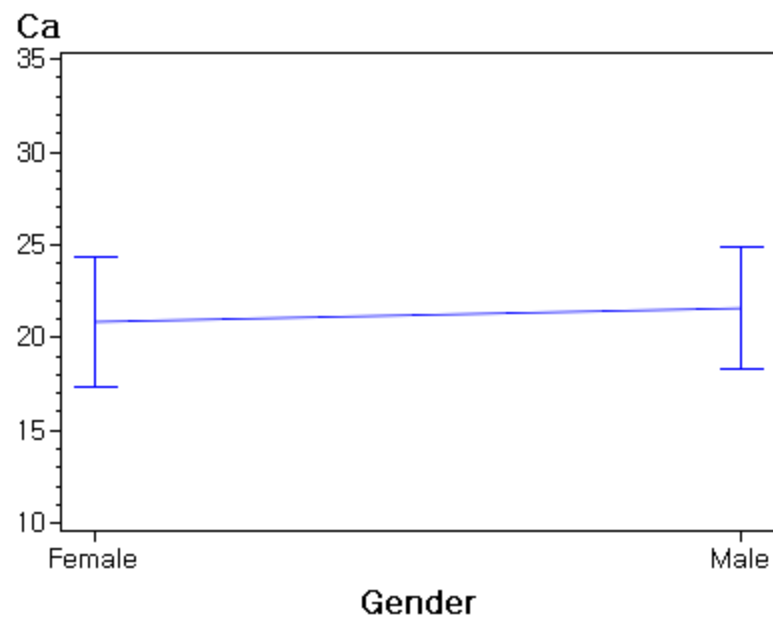
☐ Type IV(E)

☒ 모수 추정값 표시(M)

☐ 모수 추정값의 신뢰한계(L)

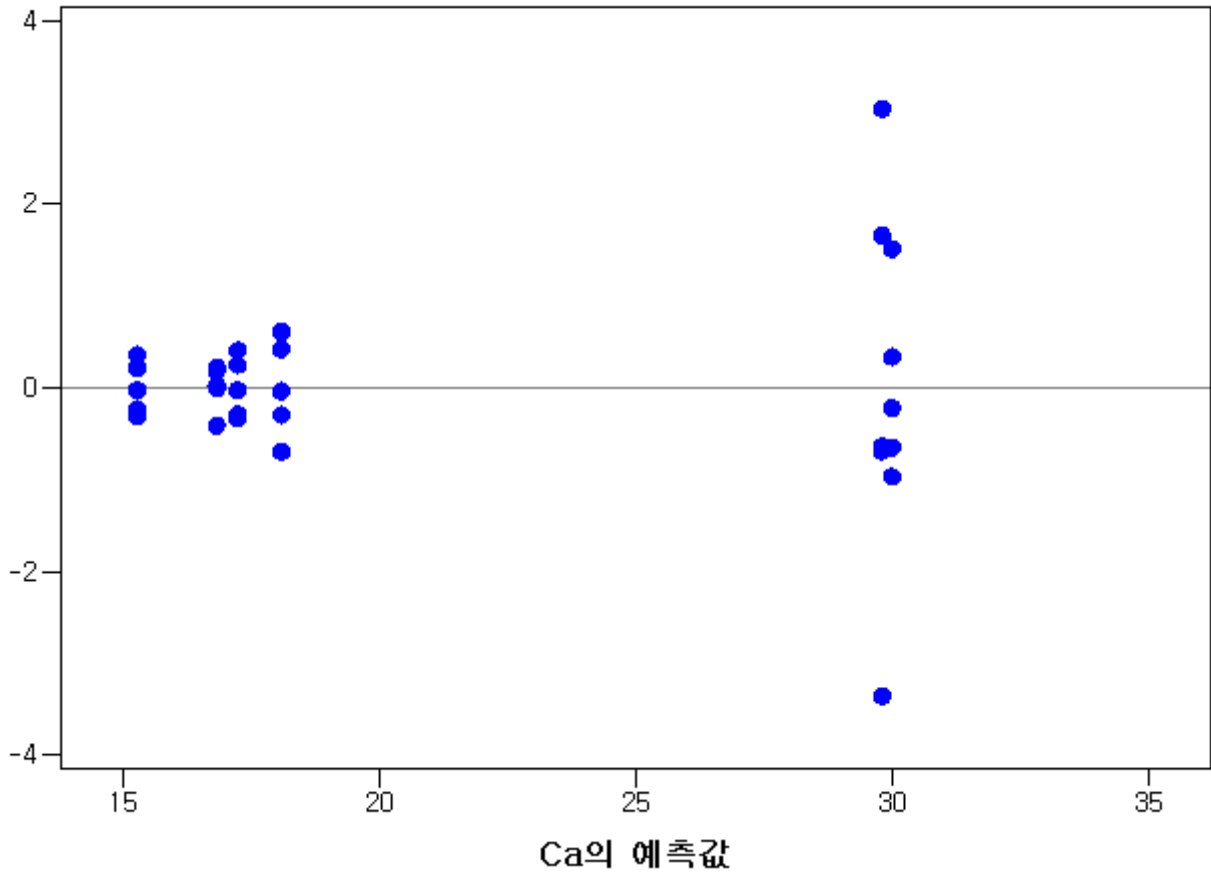
신뢰수준(V):

95%



등분산 가정에 위배

Ca의 표준화 잔차



- 선형모형 관점의 분석
 - Proc ANOVA
 - Proc GLM
- 분산분석표 : 분산분석모형에 대한 검정

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1154.550147	230.910029	72.64	<.0001
Error	24	76.292400	3.178850		
Corrected Total	29	1230.842547			

- 분산분석모형의 설명력

R-Square	Coeff Var	Root MSE	Ca Mean
0.938016	8.403983	1.782933	21.21533

- 주효과 및 교호작용 효과에 대한 분산분석표

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Gender	1	4.062720	4.062720	1.28	0.2694
Trt	2	1146.642007	573.321003	180.35	<.0001
Gender*Trt	2	3.845420	1.922710	0.60	0.5543

교호작용효과는 유의한 차이(영향)를 보여주지 못하므로 모형은 주효과로만 표현

- 제3 유형의 제곱합(SS3)

효과 통제후 남은 효과를 분석하는 제곱합

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Gender	1	4.062720	4.062720	1.28	0.2694
Trt	2	1146.642007	573.321003	180.35	<.0001
Gender*Trt	2	3.845420	1.922710	0.60	0.5543

- 분산분석모형의 계수 추정치

Parameter	Estimate		Standard Error	t Value	Pr > t
Intercept	29.81200000	B	0.79735187	37.39	<.0001
Gender Female	0.19400000	B	1.12762582	0.17	0.8648
Gender Male	0.00000000	B	.	.	.
Trt Tr1	-12.97400000	B	1.12762582	-11.51	<.0001
Trt Tr2	-11.71200000	B	1.12762582	-10.39	<.0001
Trt Tr3	0.00000000	B	.	.	.
Gender*Trt Female Tr1	-1.74200000	B	1.59470373	-1.09	0.2855
Gender*Trt Female Tr2	-1.04800000	B	1.59470373	-0.66	0.5173
Gender*Trt Female Tr3	0.00000000	B	.	.	.
Gender*Trt Male Tr1	0.00000000	B	.	.	.
Gender*Trt Male Tr2	0.00000000	B	.	.	.
Gender*Trt Male Tr3	0.00000000	B	.	.	.



교호작용효과의 계수 추정치

[참고] 모형 제곱합에 대한 분할 제곱합의 표현 방법

- . Type I SS ; Model문에서 설명요인의 표현 순서에 따라 모형 제곱합이 얼마만큼 증분되는가를 표현
 - 모형에서 요인들의 순서를 어떻게 두느냐에 따라 값이 달라짐
- . Type III SS ; 반응변수에 영향을 미치는 여러 요인의 효과를 제거(보정)한 후 각 설명요인이 고유하게 기여하는 제곱합(요인의 고유기여분)을 표현
 - 분산분석모형에서는 동일하나 공분산분석에서는 제 1종 제곱합과 다르게 나타남.

Non-parametric 2-way ANOVA

- Friedman's 2-way ANOVA
 - Assumptions of 2-way ANOVA in parametric method
 - No missing observations per cells
 - No requirement for the data to be Normally distributed, the residual are expected to have a Normal distribution
(the model is appropriate does not necessarily follow)
 - Non-parametric 2-way ANOVA
 - no fulfil the assumptions of the parametric method
 - the model will not fit well
 - wide variation in the s.d.'s

Friedman's 2-way ANOVA

- Procedures
 - give the rank for each subject
 - compute the rank-sum in the i -th group
(a similar way to the Kruscal-Wallis)
 - Calculate the statistic H
 - k groups and n subjects (no replications)

$$H = \frac{12}{nk(k+1)} \sum_1^k [R_i - n(k+1)/2]^2 \sim \chi^2(k-1) \text{ under } H_0$$

- $n(k+1)/2$ is the expected value for R_i if the null hypothesis is true and all groups are the same

$$H = \frac{12}{nk(k+1)} \sum_1^k R_i^2 - 3n(k+1) \quad \text{only one observation in each cell and a few ties}$$

Example : Immersion suit leakage

- Immersion suit leakage(g) during simulated helicopter underwater escape (Light et al, 1987)

subjects	Diving Suit type			
	A	B	C	D
1	308	132	454	64
2	102	526	0	28
3	182	134	96	30
4	268	324	264	90
5	166	228	134	34
6	332	296	458	6
7	198	350	200	90
8	28	274	16	24
Mean	198	283	202.75	45.75
SD	103.06	127.33	178.94	31.63

subjects	Diving Suit type_rank			
	A	B	C	D
1	3	2	4	1
2	3	4	1	2
3	4	3	2	1
4	3	4	2	1
5	3	4	2	1
6	3	2	4	1
7	2	4	3	1
8	3	4	1	2
sum	24	27	19	10
mean	3.00	3.38	2.38	1.25

$$H = \frac{12}{nk(k+1)} \sum_1^k R_i^2 - 3n(k+1) = 12.45$$

$$p - value = \Pr(\chi^2(3) < 12.45) < 0.01$$

- Multiple comparisons
 - Comparison of pairs of groups : Wilcoxon matched pair test
- the Friedman's test with two groups is equivalent to an extension of the sign test rather than the Wilcoxon test